Wavelet Transforms as Applied to Financial Market Analysis

Mac A. Cody
E-Systems, Garland Division
Dallas, TX

Presented at the conference

Evolving Complexity
Challenges to Society, Economy and the Individual

The University of Texas at Dallas
November 18 & 19, 1994

Abstract

The wavelet transform (WT) has been recently introduced as a tool for the analysis of time-domain functions. The WT translates functions into time-scale representations. A function is decomposed by the transform into sets of coefficients representing the function at different levels of scale with different resolutions (coarseness) in time. The coefficients generated by the WT are essentially correlations between the given function and the translated and dilated 'mother' wavelet and scaling functions used in the transform. Given certain types of wavelets, the transform coefficients can be reconstructed by the inverse WT into the original function without error.

The time-scale nature of the WT results in several useful properties. Transient events present in the function are preserved in the transformation. Non-stochastic functions can be processed by the WT without masking their nature. These properties of the WT are useful in the analysis of some economic functions, which are believed to have time-variant, multi-scale relationships.

This paper presents the concept of wavelets and the WT. Methods for generating different types of wavelets will be discussed. Two discrete-time variants of the WT, the Discrete Wavelet Transform (DWT) and the Wavelet Packet Transform will be presented. Several possible applications of the DWT to financial market analysis will be discussed.

Introduction

Within recent years, increasingly sophisticated mathematical tools have been used to analyze complex financial systems. Many of these tools were originally developed for the scientific and engineering communities and have been widely used for years. One example is the ubiquitous Fourier Transform (FT), which has become a cornerstone of modern data analysis. A function in the time domain is translated by the FT into a function in the complex (real and imaginary) frequency domain, where it can be analyzed for its frequency content. The FT describes the original function in terms of orthogonal basis functions of sines and cosines of infinite duration. The Fourier coefficients of the transformed function represent the contribution of the sines and cosines at each frequency. The FT is most commonly used in the form of the Discrete Fourier Transform (DFT), which analyzes discretely sampled time series.

The DFT works under the assumption that the original time-domain function is periodic in nature. As a result, the DFT has difficulty with functions that have transient components, that is, components which are localized in time. This is especially apparent when a signal has sharp transitions. Another problem is that the DFT of a signal does not convey any information pertaining to translation of the signal in time other than the phase of each Fourier coefficient. The phase values are averaged for the input function. Applications that use the DFT often work around the first problem by windowing the input data so that the sampled values converge to zero at the endpoints. Attempts to solve the second problem, such as the development of the short-time Fourier transform (STFT), have met with marginal success.

In recent years, new families of orthogonal basis functions have been discovered that lead to transforms which overcome the problems of the DFT. These basis functions are called wavelets, which, unlike the sine and cosine wave of the FT, do not have to have infinite duration. They can be non-zero for only a small range of the wavelet function. This "compact support" allows the wavelet transform (WT) to translate a time-domain function into a representation that is not only localized in frequency (like the FT) but in time as well. This ability has brought forth new developments in the fields of signal analysis, image processing, and data compression.

This paper will give a basic understanding of what wavelets are and how they can be used to represent signals through the Wavelet Transform (1). Descriptions of the computationally efficient discrete wavelet transform (DWT) and its generalization, the Wavelet Packet Transform (WPT), are also presented (2). Finally, there will be a discussion of how the DWT and WPT can be applied to the analysis of financial market data.

What is a Wavelet?

There are two fundamental equations upon which wavelet calculations are based. These are the scaling function (also called the dilation equation or fundamental recursion) and the primary wavelet function:

$$\Phi(t) = \sum_{k \in Z} a_k \Phi(2t - k) \qquad \Psi(t) = \sum_{k \in Z} (-1^k) a_{k+l} \Phi(2t + k),$$

where Z is the set of integers and the a_k are the wavelet coefficients. Both of these functions are two-scale difference equations and are the prototypes of a class of orthonormal basis functions of the form:

$$\Phi_{j,k}(t) = 2^{j/2}\Phi(2^j t - k);\ j,k \in Z \qquad \Psi_{j,k}(t) = 2^{j/2}\Psi(2^j t - k);\ j,k \in Z.$$

The parameter j controls the dilation or compression of the function in time scale as well as in amplitude. The parameter k controls the translation of the function in time. The set of basis functions formed by $\Phi(t)$ and $\Psi(t)$ is a system of scaled and translated wavelets.

Wavelet systems can be either real or complex-valued, though most research has used real-valued wavelet systems. Wavelet systems may or may not have compact support. As mentioned earlier, wavelets have compact support if and only if they have a finite number of non-zero coefficients. Since compact support is what gives wavelets the ability to localize in both time and frequency, this paper will be dealing only with wavelets of that type.

Several techniques have been used to create wavelet systems. These include cubic splines, complex exponentials, and parameter space constructions. Parameter space constructions have been used to construct wavelets which are orthonormal bases. The Haar wavelet (the first wavelet basis discovered) and the Daubechies wavelets belong to this class; and are the type used in this paper. Orthonormal wavelets allow perfect reconstruction of a function from its wavelet transform coefficients by the inverse WT. A set of conditions must be satisfied before a set of coefficients can represent a orthonormal wavelet. These conditions are as follows:

$$\sum_{k \in Z} a_{2k} = 1/\sqrt{2} \qquad \sum_{k \in Z} a_{2k+1} = 1/\sqrt{2} \qquad \sum_{k \in Z} a_k a_{k+2l} = 0 \ \ for\ l \neq 0 \qquad \sum_{k \in Z} \bar{a}_k a_k = 1,$$

where $\bar{a}_k$ is the complex conjugate of a_k (a_k can be complex).

Deriving and Implementing Wavelet Transforms

Once a wavelet system is created, it can be used to expand a function $g(t)$ in terms of the basis functions:

$$g(t) = \sum_{l=-\infty}^{\infty} c_l \phi_l(t) + \sum_{j=0}^{\infty} \sum_{k=-\infty}^{\infty} d_{j,k} \psi_{j,k}(t)$$

with the coefficients calculated by inner products as:

$$c(l) = <\phi_l(t) \mid g> = \int g(t)\phi_l(t)dt \qquad d(j,k) = <\psi_{j,k}(t) \mid g> = \int g(t)\psi_{j,k}(t)dt.$$

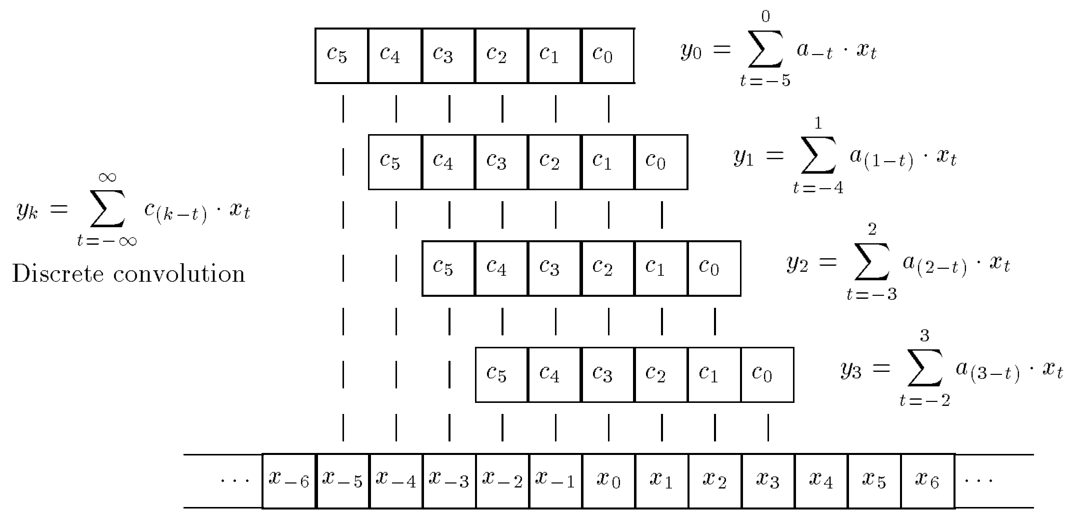


Figure 1:

If the wavelet system has compact support and an upper limit is placed upon the degree of dilation j, then the expansion equation becomes:

$$g(t) = \sum_{l \in Z} c_l \phi_l(t) + \sum_{j=0}^{J-1} \sum_{k \in Z} d_{j,k} \psi_{j,k}(t).$$

The expansion coefficients c_l represent the approximation of the original signal $g(t)$ with a resolution of one point per every 2^j points of the original signal. The expansion coefficients $d_{j,k}$ represent the details of the original signal at different levels of resolution. These coefficients completely and uniquely describe the original signal and can be used in ways similar to the Fourier transform. The Wavelet Transform, then, is the process of determining the values of c_l and $d_{j,k}$ for a given $g(t)$ and wavelet system.

The expansion equation naturally leads to a recursive algorithm for the wavelet transform if some assumptions are made. First, the function $g(t)$ is taken as a sequence of discrete points, $y_{J,p}$, sampled at 2^J points per unit interval. These points can be viewed as the inner products of $\Psi_{J,p}$ and $g(t)$. That is, the sample points are an approximation, or c_l coefficients, of the continuous function $g(t)$. This allows the c_l and $d_{j,k}$ terms to be calculated by convolution of the samples of $g(t)$ with the wavelet coefficients a_k.

Daubechies has discovered that the WT can be implemented with a specially designed pair of finite impulse response (FIR) filters called a quadriture mirror filter (QMF) pair. A FIR filter performs the dot product (or sum of products) between the filter coefficients and the discrete data samples. The act of passing a set of discrete samples, representing a signal, through a FIR filter is a discrete convolution of the signal with the coefficients of the filter (Figure 1).

The outputs of the QMF filter pair are decimated (or desampled) by a factor of two, that is, every other output sample of the filter is kept, the others being discarded. The low frequency (low-pass) filter output is fed into another identical QMF filter pair. This operation can be repeated recursively as a tree or pyramid algorithm (Figure 2), yielding a group of signals that divides the spectrum of the original signal into octave bands with successively coarser measurements in time as the width of each spectral band narrows and decreases in frequency (Figure 3).

Mallat has shown that the tree or pyramid algorithm can be applied to the wavelet transform by using the wavelet coefficients as the filter coefficients of the QMF filter pairs. The same wavelet coefficients are used in both the low-pass and high-pass (actually band-pass) filters. The low-pass filter coefficients are associated with the a_k of the scaling function Φ. The output of each low-pass filter is the c_l, or approximation components, of the original signal for that level of the tree. The high-pass filter is associated with the a_k of the wavelet function Ψ (n.b. the alternating sign change). The output of each high-pass filter is the $d_{j,k}$, or detail components, of the original signal at resolution 2^j. The c_l of the previous level are used to generate

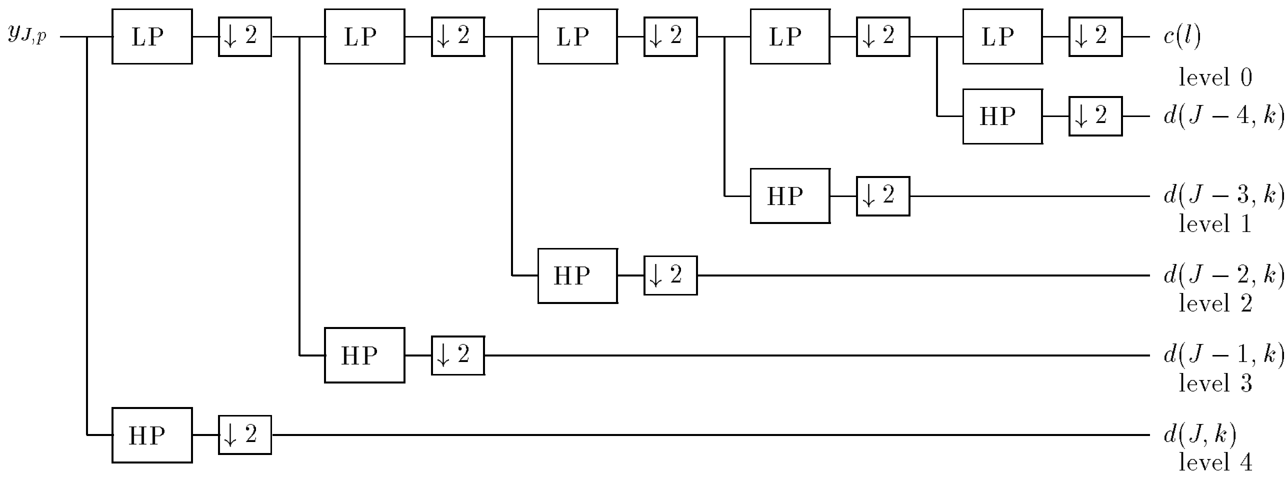


Figure 2:

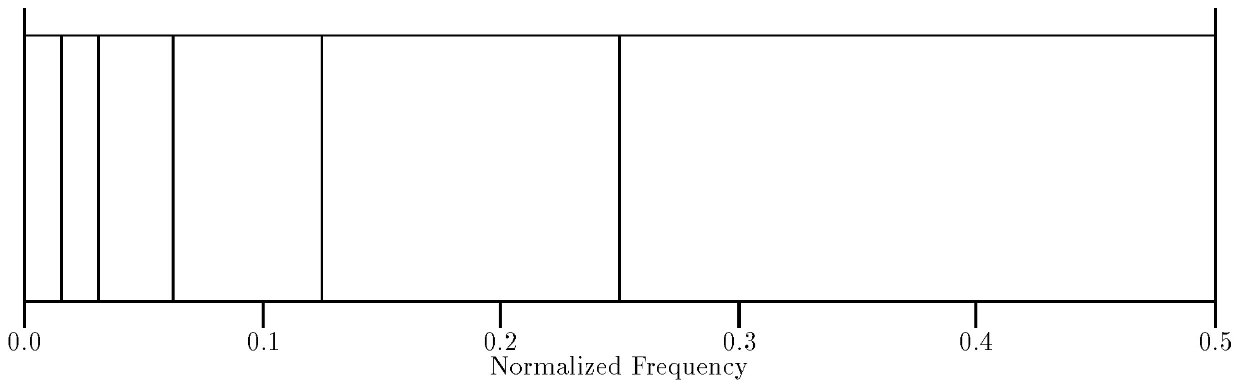


Figure 3:

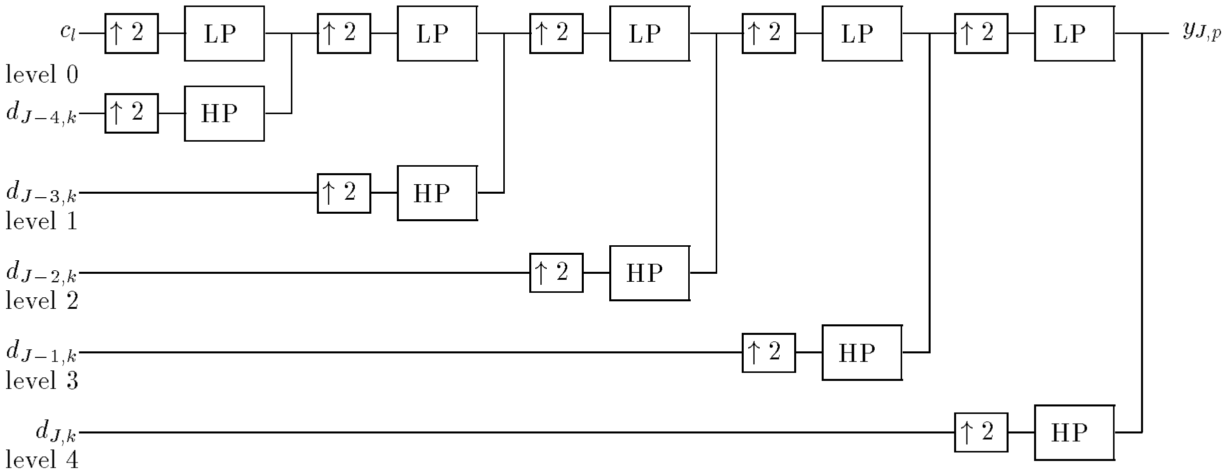


Figure 4:

the new c_l and $d_{j,k}$ for the next level of the tree. The decimation by two corresponds to the multi-resolution nature (the j parameter) of the scaling and wavelet functions.

The inverse DWT essentially performs the operations associated with the DWT in the opposite direction. The expansion coefficients are combined to reconstruct the original signal. The same a_k coefficients are used as in the forward transform, except the ordering of the coefficients is reversed. The process works down the branches of the tree combining the approximation and detail signals into approximation signals with higher levels of detail (Figure 4). Instead of decimation, the signals are interpolated, that is, zeros are placed between each approximation and detail sample and the new signals passed through the low-pass and high-pass filters, respectively. The intermediate zero values are replaced by 'estimates' derived from the convolutions. The outputs of the filters are then summed to form the approximation coefficients for the next higher level of resolution. The final set of approximation coefficients at the top level of the tree in the reverse transform is a reconstruction of the original signal data points.

An added bonus of the DWT is that it is actually more computationally efficient than the FFT. As mentioned previously, an FFT of length N (where N is an integral power of 2) takes on the order of $Nlog_2(N)$ operations. A DWT of length N requires on the order of only N operations. The FIR filters are linear processing elements. Only half as many operations are performed on each level of resolution in the tree algorithm as in the previous level.

The DWT is a subset of a far more versatile transform, the Wavelet Packet Transform (WPT). Developed by Dr. Ronald A. Coifman of Yale University, the wavelet packet transform generalizes the time-frequency analysis of the DWT. It yields a family of orthonormal transform bases of which the wavelet transform basis is but one member. The relationship between the DWT and the WPT can be most readily observed through manipulation of the tree algorithm described above.

If the tree algorithm illustration in Figure 2 is turned on end, the initial data input is at the top and the detail and approximation coefficients fan out below it. The DWT tree algorithm can now be viewed as a partial graph of a binary tree (the significance of this will be seen shortly); see Figure 5. The flow of the algorithm moves down and to the left, forming new levels of the transform from the approximation coefficients at higher levels. The detail 'branches' are not used for further calculations.

Now, the set of detail and approximation coefficients at each level of the transform form a pair of subspaces of the approximation coefficients of the next higher level of scale and, ultimately, the original data set. The subspaces created by the wavelet transform roughly correspond to the frequency subbands partitioning the frequency bandwidth of the data set, as shown in Figure 3. These subspaces, then, form a "disjoint cover" of the frequency space of the original data set. In other words, the subspaces have no elements in common and the union of the frequency subbands span the frequency span of the original data set.

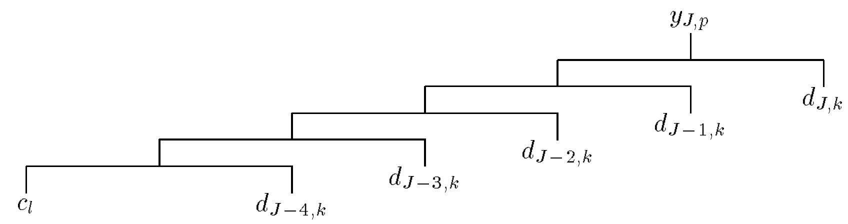


Figure 5:

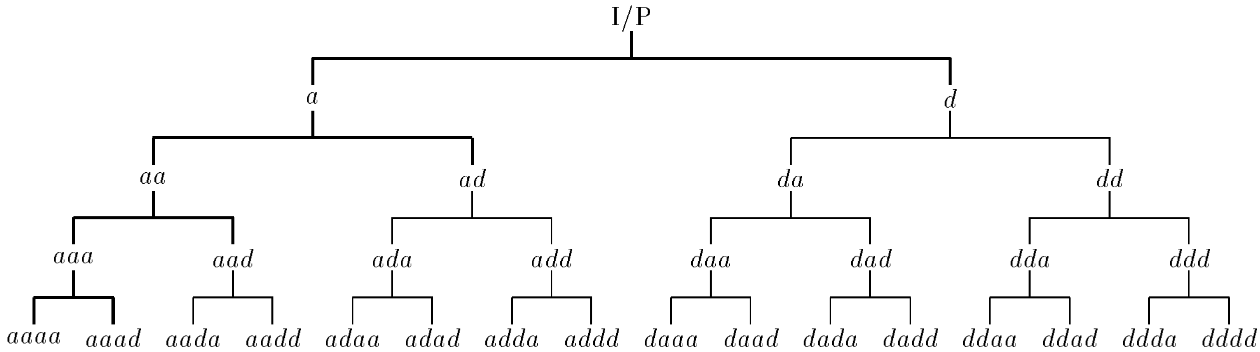


Figure 6:

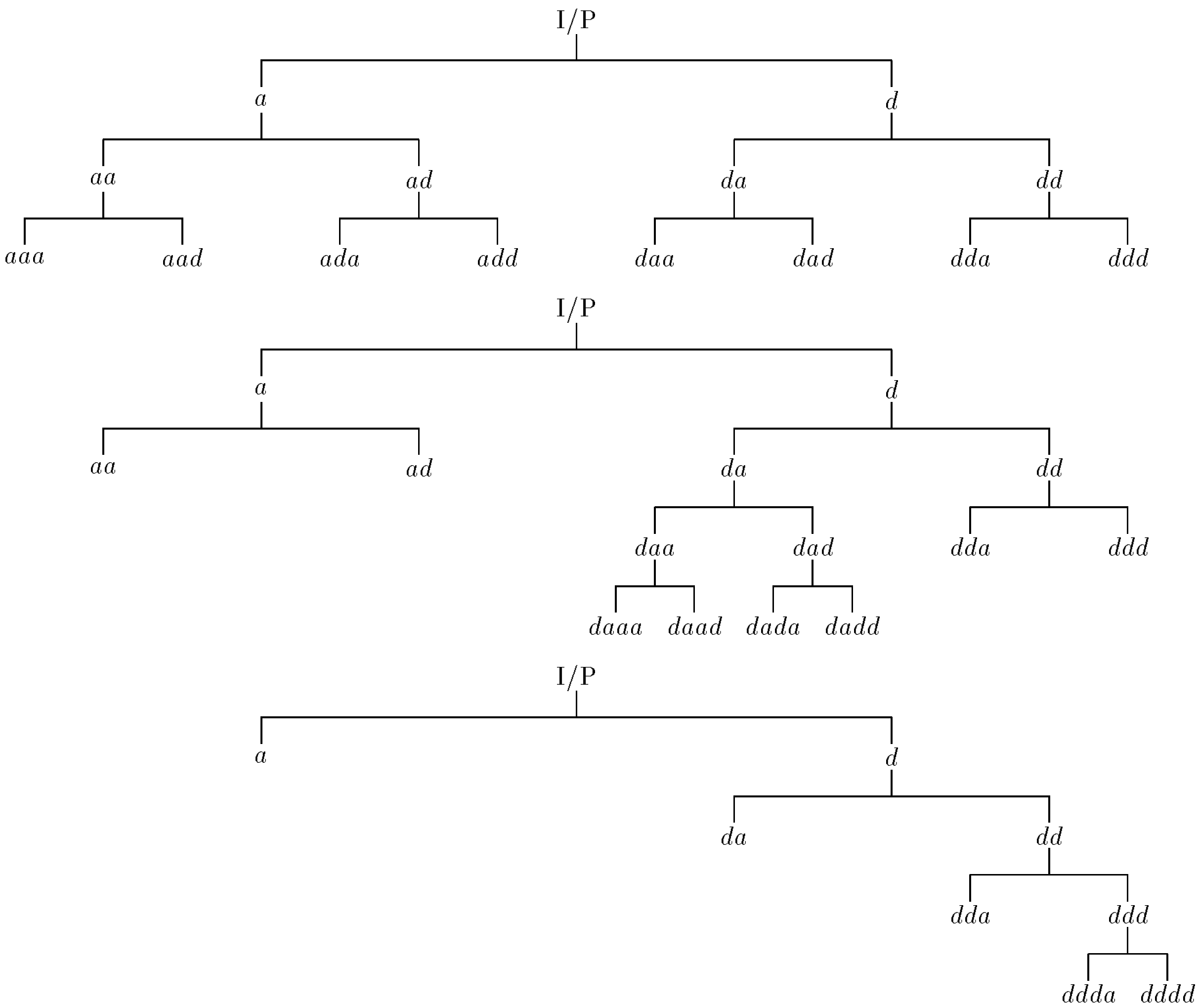
Dr. Coifman proposed that any set of subspaces which are a disjoint cover of the original data set is an orthonormal basis. The wavelet transform basis is then but one of a family of orthonormal bases with different subband intervals. As with the wavelet transform basis, each disjoint cover roughly corresponds to a covering of the frequency space of the original signal. Coifman dubbed this family a wavelet packet library. The various orthonormal bases are formed by arbitrary applications of the QMF filter pair upon the detail coefficients and the approximation coefficients of higher transform levels.

The application of the filter pair to both the detail and approximation coefficients results in an expansion of the structure of the DWT tree algorithm. The tree algorithm for the wavelet packet transform can be represented as a full binary tree; see Figure 6. The 'a' and 'd' symbols at each node, when read from left to right, indicate the order of QMF filter operations performed which yield each particular subspace of the original data set. Each node in the transform tree is also representative of a particular wavelet packet. The transform coefficients computed at each node are a correlation of the original data set and a waveform function representing the wavelet packet. Different wavelet packet bases can be formed via arbitrary "pruning" of the branches of the binary tree. A WPT with J levels yields $J * 2^J$ graphs representing different orthonormal bases; some examples are shown in Figure 7.

The variety of orthonormal bases which can be formed by the Wavelet Packet Transform algorithm, coupled with the infinite number of wavelet and scaling functions which can be created, yields a very flexible analysis tool. The WPT allows tailoring of the wavelet analysis to selectively localize spectral bands in the input data as well as correlate the signal to the wavelet. Not only can the best wavelet be chosen to analyze a particular signal but the best orthonormal basis as well. In signal processing terminology, the various bases of the WPT can be used as arbitrary adaptive tree-structured filter banks.

Applications to Financial Markets Analysis

Depending upon the "school of thought", the statistics of economic systems, such as financial market indicators, may be either non-stationary and Gaussian, stationary and non-Gaussian (3), some form of deterministic chaos/fractal behavior, or possibly a combination of these (4). The wavelet transform is positioned to enable



I/P
a
d
aa
ad
da
dd
aaa
aad
ada
add
daa
dad
dda
ddd
I/P
a
d
aa
ad
da
dd
daa
dad
dda
ddd
daaa
daad
dada
dadd
I/P
a
d
da
dd
dda
ddd
ddda
dddd

Figure 7:

one to perform analysis without assumptions being made with respect to either the stationarity of the data or the statistical distribution embodied in the data. These features are retained because the WT is a linear operator which preserves the time element of the data in the transform coefficients at the different levels of scale. The proposed applications presented below, use DWT to analyze financial market data. The WPT, with its added flexibility, could be used in a similar fashion.

Multi-scale Prediction

The first application uses the DWT as a preprocessor for a "multi-scale" linear prediction system. The concept behind this system is to use the DWT to decompose the data into "subspaces" that are easier to model with multiple linear predictors. A Daubechies wavelet is used in this application. For a given wavelet filter length, the Daubechies wavelets are optimum at concentrating the signal power into the lower-scale transform coefficients. With the significant component of the signal power at the lower levels of scale, fewer significant coefficients will need to be generated in the prediction process. As a result, there is less "risk" in the prediction. The steps for performing the multi-scale linear prediction are as follows:

1) Calculate the N-level DWT of the entire data set. Use a Daubechies wavelet to concentrate the data energy towards the approximation coefficients.

2) Generate the coefficients for the Maximum Entropy Method spectral estimation technique at each level of scale corresponding to time t_0 through t_n, which is on a DWT unit interval boundary.

3) Predict DWT transform coefficients for each level of scale for the next k unit intervals. Compare to the actual DWT coefficient values.

4) Use the inverse DWT to construct a predicted data sequence for the next k unit intervals from time t_{n+1} onward. Compare to the actual data set values.

Fractal Dimension Analysis

Another application of the DWT is to use it as an analysis tool to determine if the fractal dimension of a market indicator maintains consistency through different levels of scale. In Peters (5), he discusses that financial market indicators are non-linear dynamic systems that have chaotic/fractal characteristics. The fractal dimension of a market indicator gives insight into modeling the dynamics of the system. It may prove enlightening to see if the fractal dimension of the indicator remains consistent over different levels of scale. In addition the selection of the orthonormal wavelet used in the transform may influence the result, since these wavelets are themselves recursively defined and fractal in nature. Being a close approximation of the fractal dimension, Peters uses the correlation dimension as the metric for his analysis and will be used here. The steps in the application are as follows:

1) Determine the correlation dimension of the data set.

2) Calculate the N-level DWT of the entire data set. Various orthonormal wavelets could be used.

3) Determine the correlation dimensions of the different levels of scale (both detail and approximation data) and compare to that of the original data. Does the wavelet chosen make a difference?

4) Compare the correlation dimensions of different market indicators at different levels of scale to detect if they are influenced by similar underlying processes (similar dynamics).

De-noising of Market Indicator Data

A third application of the DWT is to de-noise market indicator data. As indicated by Peters, there is questionable value in having too fine a sampling resolution in the market indicator data. For example, daily measures of the S&P 500 contains significant "noise" component which masks any underlying statistic. As a result, most market analyses either sample the indicator at a lower resolution (e.g. monthly) or use a 30-, 60-, or 90-day moving average to remove the "noise". Techniques have been developed using the DWT to reduce the noise content in signals. Performing such an operation may reveal hidden short-term market patterns or clarify long term patterns where there is insufficient time span to cover more than a few cycles in the market indicator.

One de-noising method, developed by Mallat and Hwang (6), uses the wavelet modulus maxima as an indicator of significant "edges" in a data set. A modulus maxima is a local maximum point in a data set The absolute values of points on either side of the candidate point are less than the absolute value of that point. The wavelet modulus maxima is the local maxima for the different levels of scale of the transformed data. The steps for de-noising the market indicator data are as follows:

1) Calculate the N-level DWT of the entire data set. Use a Daubechies wavelet to concentrate the data energy towards the approximation coefficients.

2) Determine the wavelet modulus maxima in the transformed data.

3) Remove the modulus maxima whose amplitudes increase on average when the scale decreases, or which do not propagate to larger scales.

4) Reconstruct the "de-noised" data by taking the inverse DWT of the approximation level of the transformed data and the modulus maxima in the detail levels that were retained.

Feature Space Mapping of Multi-scale Data

A final application uses the DWT as a preprocessor for input into a Self-Organizing Map (SOM) neural network. The SOM, developed by Kohonen (7), is a self-training network, which maps the vector-space distribution of the input data into a matrix of interconnected nodes. After training, each node acts as a representation of a particular class of the data vectors that were input into the network. Nodes which are adjacent to one another represent similar, but not identical vector classes. This occurs due to the adaptation algorithm used in training the SOM network. The node which most closely matches the input vector has its weights adjusted. Those nodes w hich are neighbors to the "winning" node also have their weights adjusted. The size of the neighborhood surrounding the winning node starts large and decreases as the network trains.

Using the output of the DWT to form the input vectors into the SOM results in a time-scale distribution map of the data processed. Since the DWT is not time invariant, a complete map requires multiple calculations of the DWT with the data shifted across the transform unit interval. There are 2^N unique transforms for an N-level DWT. The procedure for creating the SOM is as follows:

1) Calculate 2^N, N-level sliding DWTs, with each transform translated so the N unit intervals do not fully overlap. Process the market data using a Daubechies wavelet.

2) At each unit interval of each sliding DWT, apply the unit interval as a normalized 2^Ncontains significant-point input vector into the SOM. Train until the data is exhausted.

3) After training is complete, run the same market data through the system with the trained SOM. and observe the vector trajectory over the map. Run similar market data through the trained SOM and compare the resulting trajectory to the first one.

Conclusion

The time-scale analysis afforded by the DWT/WPT provides a new approach to studying financial market data. Time-varying statistical features in the data are retained in the transform coefficients. Selection of the appropriate wavelet allows correlation of the data to the mother wavelet at different levels of scale. The applications described above suggest some ways in which the DWT/WPT can be applied.

References

1. Cody, Mac A. "The Fast Wavelet Transform." *Dr. Dobb's Journal,* April, 1992, pg. 16 - 28.

2. Cody, Mac A. "The Wavelet Packet Transform." *Dr. Dobb's Journal,* April, 1994, pg. 44 - 54.

3. Mandelbrot, Benoit B., *The Fractal Geometry of Nature,* W. H. Freeman and Company, 1983, ISBN 0-7167-1186-9, pg. 334 - 340.

4. Dooley, Kevin J., Timothy L. Johnson, and David H. Bush, *Some Relationships Between Chaos, Randomness, and Statistics.*

5. Peters, Edgar E., *Chaos and Order in the Capital Markets,* 1991, John Wiley & Sons, Inc., ISBN 0-471-53372-6.

6. Mallat, Stephane, and Wen Liang Hwang, "Singularity Detection and Processing with Wavelets", *IEEE Transactions on Information Theory,* Vol. 38, No, 2, March 1992, pg. 617 - 643.

7. Kohonen, Teuvo, *Self-Organization and Associative Memory,* Second Edition, 1988, Springer-Verlag, ISBN 3-540-18314-0.